DOCUMENT RESUME

ED 223 716 TM 820 863

AUTHOR	Sweeney, Jim; Manatt, Richard
TITLE	Teacher Competence: The Past, Present and Future of Its Assessment.
PUB DATE	[82]
NOTE	38p.
PUB TYPE	Information Analyses (070) -- Reports - Evaluative/Feasibility (142)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	Adults; *Competence; Evaluation Criteria; *Evaluation Methods; Literature Reviews; Models; *Teacher Evaluation; Teacher Qualifications; *Teachers
IDENTIFIERS	*Redfern Approach; *Teacher Performance Evaluation (Manatt)

ABSTRACT

This article highlights the state of the art in assessing the competency of elementary and secondary teachers in America. It includes an analysis of the historical perspective of teacher evaluation, the dilemma currently plaguing teacher evaluators, and future direction for research in teacher evaluation. The primary focus is on the two models which are considered the most efficacious for evaluating teachers: the Redfern Model and Teacher Performance Evaluation. Both models are explicated and a discussion of strengths and weaknesses provided. States requiring teacher evaluation and evaluation system criteria are included in the appendices. (Author/PN)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED223716

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

Teacher Competence: The Past Present and Future of Its Assessment

By

Dr. Jim Sweeney

and

Dr. Richard Manatt

Iowa State University

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
J. Sweeney & R. Manatt
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM 820 863

Teacher Competence: The Past, Present, and Future of Its Assessment

If education is the cornerstone upon which a great nation builds, then teaching must be one of our most important human activities. Assessing the competence or proficiency of the more than two million public elementary and secondary teachers who shoulder the responsibility for educating America's youth has become a hot issue, triggering a proliferation of journal articles and books replete with schemes and tips on how to evaluate teachers. The concern is hardly surprising; a 1978 study by Educational Research Services found that 98 percent of the districts surveyed have some type of formal evaluation of teachers. While teacher performance evaluation efforts have been applauded publicly, there are those who question its efficacy. Scriven, of the Institute of Evaluation at the University of San Francisco, describes teacher evaluation as a disaster, where practices are shoddy and principles unclear (Millman, 1981). He maintains that while recent work has identified some ways to clarify the issues and make procedures equitable and reasonably valid, there is not a single exemplary system in which the practice comes near to matching the knowledge. A paper on the subject seems timely, for we must desist from feverishly building better mouse-traps and examine the state of the art: where we've been, where we are, and where we wish to go.

As we developed this paper we were reminded of our youthful experiences with erector sets. What was thought to be the finishing touches on a windmill sometimes turned out to be the first steps in dismantling, accompanied by the joy (and sorrow) of re-examining dozens of odd-shaped components. We gleefully set out to describe the major teacher evaluation models. There was little reason to assume that they would be complex or difficult to describe. We ran into one small problem; there were only two. What we found were a basket full of thingamabobs. While an optimist might describe teacher evaluation as chaotic, our cynical inclination forces us to be less charitable. Practice is a mess!

But our goal in this paper is not to debunk; description, explanation and suggestion have more merit.

Teacher evaluation has an etiology and a context. We've briefly addressed both in the first section. In the second section we've highlighted what we see as the major philosophical and theoretical issues confounding educators. The meat of the "current approaches" section is a description and critique of the Redfern and Teacher Performance Evaluation models. We highlight the latter since it is the direction in which we feel teacher evaluation systems should be heading. In the final section, "futures," we use our clairvoyant powers to suggest directions for future research and development.

It is instructional to further define the scope and direction of the paper. First, it is our intent to discuss teacher performance rather than an indirect measure of ability or potential. We have omitted a discussion of competency testing since the national thrust has been toward preservice testing of teachers. Next, there is the use of the term "competency". While competency has become somewhat synonymous with a hypothetical minimum standard, that does not accurately describe the concept to be reviewed in this paper. We will direct our analysis and commentary on school organizations' assessments of the level of proficiency at which teachers perform since we wish to include evaluation efforts aimed at identifying superior or excellent teaching. Finally, there is the matter of terminology. The title highlights assessment rather than evaluation because it is less restrictive. However, in the real world, teacher performance assessment customarily falls under the rubric evaluation. While we shall use these terms interchangeably, it is with some misgivings. Finally, while the paper is intended to be a scholarly examination of practices, our analysis is not value free. It reflects the biases imposed by years of experience and study in the area. For this, we make no apology.

Perspective

The assessment of teacher performance, like many activities, is cathected by past practices, by movements in government, industry, and research, and by other sociological factors influencing the values, sentiments, and preferences of American society. While the evolution of past practices is somewhat indiscernible, the scrutiny of teachers appears to have been given a coherent thrust in the English grammar school where teacher competency was operationally defined as the teacher's proficiency in classroom and student management. A series of pamphlets published by Hoole (1659, reproduced in 1668) captures the prevailing view of the era. It reveals that the burden for learning was placed on the pupil, not the teacher. If the classroom was correctly managed, then the pupil had ample opportunity for learning. A deficiency of learning was a sign of pupil irresponsibility. Providing the proper learning environment was the teacher's primary task. Thus, it became the primary criterion by which to assess competency. Early American educators followed the lead of their colleagues across the sea. Horace Mann, upon visiting common schools in Massachusetts, observed that teachers were largely responsible for managing work and spent little time with individual students. Educators assumed that all children were equally capable of learning; ergo, differences in student achievement were attributable to student motivation and the extent to which they applied themselves.

The first "assessors" on the American educational scene appeared in the early part of the eighteenth century in the form of committees of citizens whose primary tasks were to inspect the plant and check on pupil achievement. Since the citizens were lay persons, with no special professional expertise, they simply acted as checkers, making no effort to improve teaching or teachers (Wiles and Lovell, 1975).

The teacher's role and function, however, changed dramatically in the nineteenth century. While a number of factors contributed, the demise of the

one-room school and the introduction of the Prussian system of dividing children into grades by age groups were most salient. Since teachers had smaller classes and thus had more uniformity in their assignments, it was presumed that they could improve learning through knowledge of subject matter and their ability to impart learning. Dovetailing with changes on the educational scene were dramatic events in American society. Transition from an agricultural and craft society to an industrial society spawned a revolution marked by scientific management and a seemingly unquenchable thirst for efficiency. The pioneer, blazing the trail to efficiency and effectiveness was Frederick Winslow Taylor. By virtue of his success with Bethlehem Steel Company and other industries, and his classic, The Principles of Scientific Management (1911), Taylor convinced leadership personnel across the nation that specialization, standardization, and other principles of industrial engineering unlocked the gates to organizational success. Educators jumped on the bandwagon. The work of Charles Bobbitt (1912) and Elwood Cubberly (1916) were particularly noteworthy for both persuasively urged school administrators to work toward turning out a standard product with scientific measurement of the product. Children were "tabula rasa", raw material to be shaped according to the specifications of society.

Measurement, particularly as it related to efficiency, was a major concern of leaders in the private sector. Not surprisingly, it spilled over into the public sector. Two major studies reported by Davis (1964) in Contemporary Research on Teacher Effectiveness are illuminating. Boyce (1915) surveyed 350 cities of over 10,000 population, requesting a statement of their methods of determining and recording teacher efficiency. About 60 percent of the 242 cities which returned questionnaires reported using some type of rating or formal evaluation of teachers. Although the number of items on which teaching efficiency was judged ranged from as few as two items to as many as eighty, four types of

analysis prevailed: (1) descriptive reports dealing with specific points; (2) lists of questions answered by yes or no; (3) lists of items to be evaluated by a statistical classification, for example, excellent, good, medium, unsatisfactory; and (4) lists of items to each of which was assigned a definite numerical value representing the maximum score that might be given. Boyce's analysis of the instruments revealed that "discipline" was the most frequent quality assessed (98%) followed by "instructional skill" and "cooperation and loyalty" (each 60%).

King (1925), in a report to the National Education Association, reported that at least three-fourths of the large city school systems were using some type of efficiency rating. His analysis of the instruments used in these schools revealed that "technique of instruction", "personality", and "classroom management", appeared most frequently (in the order presented), along with criteria such as "professional attitude", "cooperation", "health or vitality", and others. It should be noted that "discipline" and "classroom management" were separate categories in the analysis. When aggregated, they occupied the most frequently used category. King found that single words or descriptive phrases were by far the most popular response mode followed by rankings, letters, and weighted point scales.

A singular event, the Hawthorne Studies, provided the impetus which pushed scientific management into the background, and stimulated what has been labelled, the Human Relations Era. The work of Mayo (1937) and Roethlisberger (1939), at the Hawthorne Plant of Western Electric, convinced public and private sector managers that human relations was a more salient factor in producing results than rigid adherence to scientific principles. The informal organization and interpersonal relations were found to be more important than specialization and standardization. Teacher evaluation in that period reflects that theme. In a 1945 study, Reavis and Cooper found the three items most frequently used for

rating teachers were "social relations", "personal characteristics", and "noninstructional school service". The researchers also noted that a five-point scale was most prevalent in rating teachers and that the same weaknesses that appeared in early evaluation systems were still present: namely, ambiguous terms, lack of definition of items, and items that linked two independent elements to form a single judgment.

The lines of demarcation from the human relations movement to the present are somewhat blurred, as is the progression of evaluation. The 1960's and early 1970's were punctuated by social upheaval, a search for relevance in the classroom, and a thirst for individuality and human dignity in all activities. It's not surprising that the emotional content of lessons, measured by an instrument such as the Flanders Interaction Analysis (1960) was an object of concern and scrutiny. Self-evaluation became popular; evaluation by an administrator was often a ceremonial rite, an opportunity for the supervisor to bestow general thanks and hand out other platitudes. While relevancy and humanism are still alive, they may be breathing their last gasps, suffocated by a reincarnation of scientific management, this time in the guise of a new rubric, accountability. During the past three decades, affluence, a predeliction for individuality, and a belief that every American is entitled to a better life has stimulated a dream for a richer life. The great American dream is menaced, however, by inflation, unemployment, and scarce resources. Schools are also threatened by declining enrollment, access to information, declining test scores, and parents concerned about the alleged illiteracy of high school graduates. But despite the uptight economy, it is a dynamic society and the public is demanding "a bang for their buck". While educators have been plagued by many fads, teacher evaluation appears to be here to stay for more then half of the states have enacted legislation mandating teacher evaluation.[1]

---

[1]See Appendix A for a more complete derivation of states and evaluation purposes.

Dilemmas

A discussion of the prevailing approaches to teacher assessment is inappropriate, perhaps even irrelevant, without a preface. The appropriateness of teacher evaluation is contingent upon the soundness of its underlying philosophical assumptions as well as the elegance of its theoretical components. Both, unfortunately, are bounded by an abundance of opinion, a dearth of facts, and a handful of dilemmas which emanate from the basic questions which have plagued evaluators for decades: why, who, what, and how? A cursory review of the issues and their current standing is illuminating, for they are inextricably woven into the fabric of teacher assessment, present and future.

Why?

Teacher evaluation has multiple, conflicting purposes. Not only are individual school organizations unable to agree on purpose, there is disagreement between the primary participants, teachers and principals. In the real world of schools, the purposes of evaluation are normally lumped under two complementary but dichotomous concepts: formative and summative evaluation. The former includes activities designed to gather specific, discrete, and limited data to aid in identifying and diagnosing teachers' strengths and weaknesses. It is continuous, part of a year-long process in which the supervisor, usually the building principal, observes a teacher in the classroom and gives him or her feedback on performance. While the feedback may include teacher activities other than those directly observed by the supervisor, the current trend is toward providing the teacher with specific, descriptive feedback about a recent lesson which the supervisor observed. In summative evaluation, the supervisor, to make administrative decisions, utilizes a broad sample of information about a teacher's overall performance. It is primarily designed to identify the level of competency at which the teacher performed during the evaluation period. Therein lies the rub, and the classic dilemma plaguing evaluators--they are expected to

do something *for* teachers, while implicitly they may have to do something *to* them. Teachers have developed strategies for dealing with this problem, subtle gameplaying. Thus, evaluation charades are enacted daily in school principal's offices across the nation. Teachers role play their way through conferences thinking, "No! No! No!", but saying, "Yes! Yes! Yes!".

Blumberg (1980), a noted expert, presents another gameplaying scenario.

> "Each side is aware of the game and it frequently gets played with each party feeling that he won. . .the teacher knows that he is going to be observed and evaluated. He tells his students. If they like the teacher they take care of him, and the supervisor, who probably went through the same process as a beginning teacher, observes a good lesson. The teacher gets what he wants, and the supervisor gets what he wants."

Blumberg maintains that as long as supervisors are saddled with the responsibility of evaluating teachers, they are incapable of establishing a collegial relationship with them, and therefore, are unable to provide help.

Who?

At first blush, the question seems moot. In more than 90 percent of our schools the principal has primary responsibility for formally evaluating teachers (ERS, 1978). Yet, research on the efficiency of supervision does little to support the present mode. Findings by Blumberg (1970) and others indicate that the evaluator is a source of anxiety and is, therefore, ineffective as a supervisor or helper. Other approaches are being championed. A 1974 National School Public Relations Association report indicated there was a trend toward peer evaluation emanating from the apparent ability of peers to reduce threat and open communication between the appraiser and the appraisee. But lack of training, time, and the reluctance of some teachers may hinder the movement. Self-evaluation has supporters and critics. The former contend that the teacher is the best source of information regarding competency while others hold self-evaluation is distorted by egocentrism or a tendency toward professional or personal defensiveness. In practice, its use is

limited. One-third of the school districts surveyed by ERS (1978) required teachers to evaluate themselves while other districts suggest it to teachers as an option and encourage them to share the results with the principal. There is also the matter of student input into evaluation. While according to ERS (1978) only a small percentage of schools use this methodology. There are many who advocate more extensive usage of student input. Aleamani (millman, 1981) has written persuasively on the subject. While his remarks were directed at those in higher education, they seem applicable to public schools, particularly at the secondary level. Student data, carefully interpreted and prudently used, may have some value.

Finally, there are still those who contend that output, not input, is the sole criterion of teacher competence, that evaluation should measure what students have learned or how they have changed over a period of time in working with a teacher. Popham, McNeil and others have toiled diligently in an attempt to develop a workable design to link level of teacher competency to student outcomes. The efforts have largely been in vain. The use of standardized tests to equate outcome with performance appears to be unsatisfactory. Glass (1974) strikes the issue squarely, citing the validity of the tests, their unreliability or instability from year-to-year and the dangers in the use of data taken from nonrandomly constituted classes as obstacles difficult to overcome in making valid judgments. Popham (1973) agrees, but suggests that when it is necessary to differentiate among teachers, the use of performance tests or teaching power tests, where a number of teachers are given one or more identical objectives and a sample of the measure based on the objective(s) to be administered to pupils following instruction. Group of learners can be randomly selected, assigned to be teachers, instruction given, a test inferred from the objectives administered and results assessed. The procedure may have merit. While some schools may

eventually adopt this or similar strategies, it is unlikely to attract many followers. For now, it seems more indirect measures will serve as the sources of evidence for teacher competency. The principal is presently the primary evaluator, sometimes with help from supervisory staff. Who else may become involved is indeterminate.

What?

Identifying valid criteria, has been perhaps the most perplexing of the dilemmas. Research on teaching has been conducted for many years and has generated more than 10,000 published studies (Dunkin and Biddle, 1975). Early studies were fruitless since they failed to focus on the actual process of teaching in the classroom. Recent studies and meta-analyses of research on teaching have been more productive. A relatively new criterion (or set of criteria) has been eagerly embraced by educators. Researchers have identified effective teaching behaviors through *in situ* --or in classroom--observation involving the systematic recording of a variety of specific, discrete teacher behaviors that are assumed, according to a theoretical framework, to be related to pupil growth. After these behaviors are observed and recorded, they are correlated with indices of pupil change to validate their importance. Typically, the correlational data are used to identify "more effective" and "less effective" teachers according to, for example, achievement gains made by their pupils over a semester or school year. Predictably, educators have concluded that the search for the Holy Grail has ended. In workshops across the country, evaluators are boning up on how to identify these wonderful behaviors (or lack of them) in the classroom. The identification of effective teaching behaviors has obviously been a very promising development but the same researchers who uncovered them have issued strong caveats related to the context in which they occurred. Content, grade level, and student characteristics are but a few of the factors which make generalization a risky business. Their pleas, for the most part, have fallen on deaf ears. Not only are evaluators eager to learn the recipe for good teaching, but the teachers themselves, perhaps grasping

for straws, have also endorsed the concept. It's our guess that the concept will become the primary criterion for the majority of school organizations.

How?

"How" is, of course, the major question. While there are a number of ways to skin the proverbial cat and divergent views accompanying each, perhaps that which is most important and least agreed upon is how to gather and record the evidence. Bolton (1973) summarized the sources for ascertaining the quality of a teacher's performance as follows: (1) in-class behavior of the teacher--as perceived by administrators and others, (2) out-of-classroom behavior--as perceived by the administrators and others, and (3) student accomplishments--as measured by standardized tests, teacher-made tests, student projects, and observation of student's behavior. Selecting the appropriate method becomes a matter of preference, considering purpose, cost, available time, relevance, validity, reliability and east of administration.

If there is one area in which the parties to teacher assessment agree, it is on what might be labelled, the data imperative; the necessity for basing the evaluation on reliable and valid data. Common practice has changed dramatically. Evaluators once relied heavily on what they saw as they roamed the building, coupled with then unannounced visits and how they *felt* about the teachers. Contractual agreements have rendered that practice obsolete while advances in the technology have spawned an array of systematic evidence-gathering techniques designed to minimize the influence of observer bias and maximize the reliability of the data.

There are few evaluation systems which do not culminate in a cumulative rating of the teacher. Hopefully, the appraisal is made after, not in conjunction with, classroom observation and other data collection methods, summarizing the frequency and quality of many behaviors. Since the level of proficiency is an object of concern, most instruments use some type of rating scale or a checklist which has

the characteristics or a rating scale. The Semantic Differential Scale, the Guttman Scale, and various others enjoy some popularity, but summated ratings or Likert scales appear on more evaluation instruments than do others (Borich, 1981). Unfortunately, the use of rating scales provides further indication that the science of measurement lags behind the art of evaluation. Since the distance between response alternatives on scales are often unequal, the raters (or evaluators) have a tendency to make judgments toward the end of the scale since the bottom end of the scale denotes subpar performance or incompetence. Hence, the evaluator tends to use the top end, often ignoring the data. This difficulty is even further exacerbated by a tendency to use other teachers' ratings as an anchoring point.

There is also the problem with what might be called, "grade inflation". Most teachers (and administrators) have become conditioned to receiving high grades through their public and post-secondary education. Most received an "A" in student teaching with accompanying plaudits. They tend to equate the rating scale with letter grades. As a result, a rating in the middle of the scale is viewed as "C" even though the response mode may be "meets district standards". Many administrators, well aware of this and anxious to please their teachers, award the teacher a rating toward the upper end of the scale. Some evaluators still use a checklist for evaluating teachers. While the checklist does direct the evaluator's attention and lend a degree of objectivity, Griffiths (1973) contends that it tends to make the evaluation a mechanical, routine procedure, neglects the relative weight of the items, delays with superficial items and influences evaluators to make judgments without reflection and careful analysis.

The trend appears to be away from highly formalized instruments toward narrative reporting, gathering descriptive, non-judgmental data to reproduce a "mirror of the classroom" for teachers. Acheson and Gall (1951) summarized the techniques nicely, providing a comprehensive, clear description of how-to-do-it. It includes a discussion of anecdotal record keeping, time-on-task charting and

student-teacher interaction analysis. The rationale for gathering descriptive data is simple and intuitively appealing--providing teachers with a descriptive account of the significant teacher and learner behaviors which occurred during a lesson provides an opportunity for the teacher and supervisor to examine the lesson and discuss effective teaching in a non-threatening way. Brandt (1973) agreed that this is a highly productive technique but cautions that it will demand well developed skills and dedication.

What then do we have? We have an important activity beset by conflicting purposes generating political infighting on the verge of open warfare. We have uncertainity as to who should gather evidence on elusive criteria on which there has been more agreement than disagreement, plus a tangled web of measurement devices. We, also, have some significant trends. The building principal has been accorded the responsibility of assessing teacher competence. It appears that affective teaching behaviors will become the criteria for assessing that competency. While principals struggle with the summative vs. formative dilemma, it appears that they will use narrative reporting for formative data gathering and rating scales to record the summative judgment. How evaluations will be conducted remains to be seen.

Current Approaches

School organizations employ virtually hundreds of approaches to assessing teacher competency. Most are very similar, some are remarkably different. Consider the following. As we examined the various methods, we found a school district which had a formal evaluation system with no summative component. Evaluation was designed solely to help teachers improve; therefore, a judgment on the level of proficiency was studiously avoided. In another rather large district, a much different approach was noted. The evaluation process was specifically designed to weed out unsatisfactory teachers. Any member of the school community--parent, employee, student--can request, in writing a "review of service". If the teacher

is deemed in need of remediation, a team provides intensive assistance and performance is monitored. If the teacher has not improved after five months, he or she is terminated with 30 days notice.

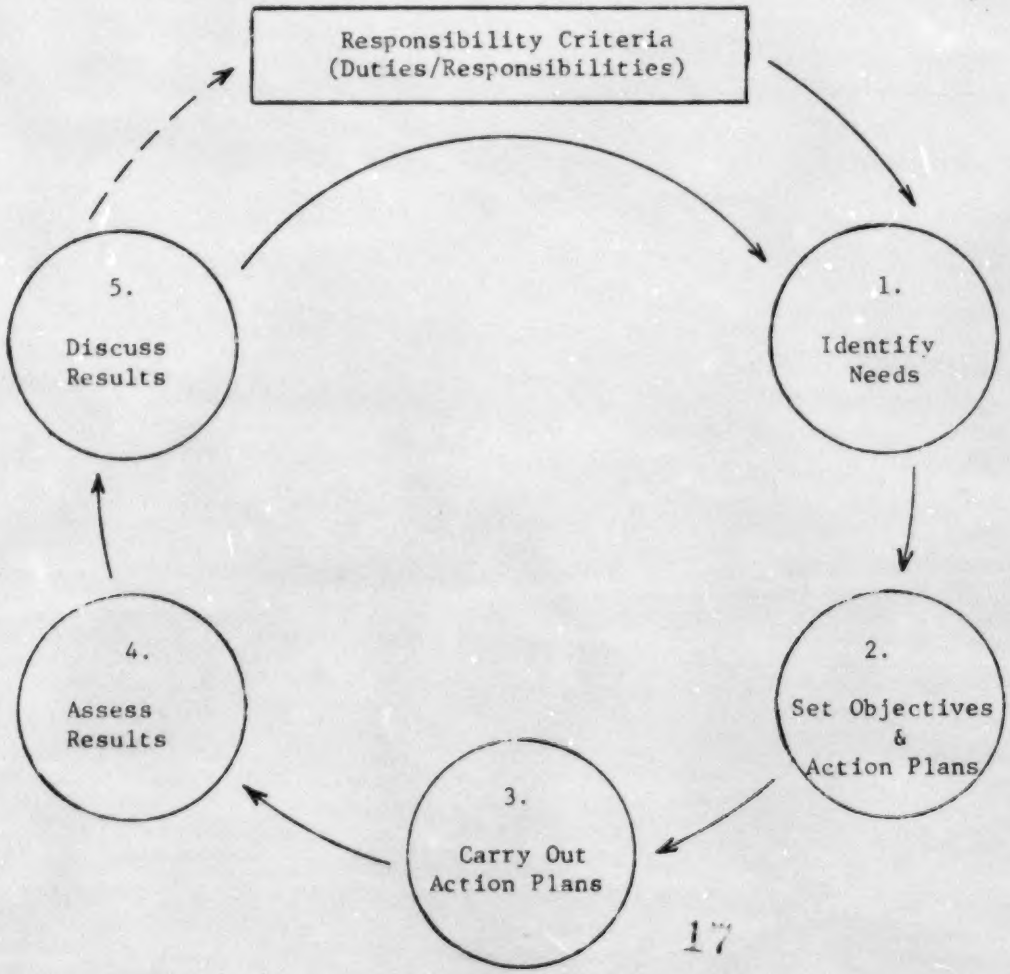
In our attempt to uncover the significant models used by teacher evaluators, we decided to forsake the product-process approaches and apply five general and 32 specific criteria. The five general criteria are those suggested by Strike and Bull (Millman, 1981) who contend that, for a school organization to make evaluations legal and morally acceptable, it must exhibit the following: (1) inclusion in the formal administrative policies of the school board, plus a reasonably precise explanation of the criteria for effective teaching and dissemination of uniform procedures for making personnel decisions; (2) a guarantee that the evaluation will focus only on those aspects of a teacher's performance behavior and activities that are directly or indirectly relevant to the teacher's ability to execute the legitimate responsibilities that attach to the job; (3) evaluation procedures that provide teachers legal due process; (4) a system that is non-alienating, produces cooperative working relationships and is aimed at increasing the professional skills of the teaching staff; and (5) provisions for sharing the evaluation data with teachers and for providing them with assistance. For the specific criteria, we applied the 32 questions which can be seen in Appendix B. A number of evaluation schemes exhibited salient characteristics, but only two, the Redfern Model and Teacher Performance Evaluation (TPE) appeared to have a sound philosophical and theoretical base and incorporate the criteria into a coherent system.

The Redfern Model

The use of behavior patterns in setting goals became widely known in the 1950's and 1960's through Peter Drucker's management-by-objectives approach in industry. This approach spread to government and in the early 1960's emerged in

schools. Essentially it asked: What are employees going to do and when are they going to do it? The Redfern Model is anchored to two assumptions: (1) individuals have inherent competencies that can be identified and described, and (2) competencies may be expressed in terms of techniques in carrying out the teaching/learning process and the attainment of measurable learner outcomes (Redfern, 1980). If performance competencies can be identified, appraisal procedures can be designed to ascertain the extent to which an individual demonstrates such competencies. The evaluator and evaluatee can work together, as partners, to reach these objectives. While the model revolves around setting and assessing performance objectives, it has six components which are sequenced to produce the cyclical process shown in Figure 1.

Figure 1.



Responsibility Criteria

The first step in the evaluation process is to define the content and scope of the job responsibility criteria. These criteria are expressed in broad areas with descriptions under each. For example, "motivating children" serves as a broad area, applicable to all teachers. A descriptor in that area might be "motivates by positive feedback, praise, and rewards". Descriptors and broad areas are linked by a numbering system which subsequently facilitates the identification of needs.

Identify Needs

Implementing this component requires a mutual effort by the evaluator and evaluatee--to identify needs, as represented by gaps between ideal and actual performance. It is equally important to focus upon areas of strength in which further gains can be made. Redfern contends that a useful way to identify needs is to regard them as areas to emphasize in order to attain the maximum degree of improvement in performance.

Set Objectives and Action Plans

Objective-setting helps to ameliorate deficiencies and meet the needs of personnel who desire improvement. While Redfern leaves the door open for those who wish to write performance objectives in general terms, the valence of specific objectives written in behavioral terms is emphasized. Objectives, whether general or behavioral, should be accompanied by an action plan--a description of how the evaluator and evaluatee will work together to reach objectives. It typically includes (1) the kinds and amount of assistance the evaluator will provide, (2) the extent of observations or other contacts that the evaluator will make with the evaluatee, (3) the frequency of follow-up and feedback conferences, and (4) an explanation of the type of data that will be collected and used in making the assessments.

Carry Out Action Plans

This, in essence, is the phase of the process where the evaluator monitors the evaluatee's performance and gathers a substantial volume of appropriate data. While collecting data, the evaluator must be sure to (1) collect relevant information, (2) obtain representative data, i.e., samplings of the full range of performance behavior, and (3) use a variety of techniques throughout the process. The evaluator works hard to assure the person being evaluated that data is being collected for analysis and that subsequently he or she (the evaluatee) will play an active role in the data analysis and assessment.

Assessment of Results

While the assessment of results focuses primarily on the extent to which performance objectives have been achieved, it does not preclude assessments of overall accomplishment. Whatever the intent, it should be clearly understood by participants prior to the process and reinforced throughout. While data may be gathered by surveying students, electronics devices (audio and video tape recording) and other sources, Redfern advocates self-evaluation and evaluator assessment as the major interacting components in the process. Self-evaluation is a guide for planning further self-improvement; evaluator assessment provides the evaluatee with objective evidence of his or her performance in the important criteria identified earlier. Assessment and data gathering are not a one-shot activity, they are continuous and ongoing. To assess the achievement of performance objectives and to assess overall performance, it is necessary to record the assessments on a Summative Evaluation Report. This report provides the evaluator an opportunity to furnish the evaluatee with comments explaining ratings or perspectives of performance. Less than satisfactory ratings are explained and the evaluatee is provided with an opportunity to make comments on the Summative Evaluation Report.

Discuss Results

According to Redfern, the evaluation conference is perhaps the most important part of the process, for it provides both parties with the opportunity to discuss progress and increase their understanding. Emphasis in evaluation conferences should not be solely on discussing problems. To increase understanding Redfern recommends that the conference discussion include: (1) discussion of long-and-short range goals and objectives, (2) recognition of good work, (3) mutual exchange of suggestions for improvement, (4) selection of top-priority job tasks or job targets, (5) clarification of the responsibilities of both parties, (6) correction of misinformation and misunderstanding, and (7) attention to whatever other topics may seem important to either party at that particular time.

The final step in the process provides the impetus for recycling the process. The evaluator and evaluatee once again identify needs based on a mutual assessment of performance based on the data and the perceptions which both bring to the assessment conference. These needs are translated into performance objectives as the cycle continues to flow. Follow-up planning by the evaluator plays an important role for it helps to outline the activities that he or she should carry out to promote the development of the individual.

Bouquets and Brickbats

The Redfern Model is a systematic approach to assessing teacher competency. It offers the following advantages:

1. Clearer perceptions of performance expectations plus evaluation based on predetermined criteria related to the employee's duties and responsibilities.
2. An opportunity for mutual discussion of employee's strengths and weaknessess plus feedback to refine performance.
3. Provides for specific, concise and measurable objectives for performance improvement.
4. Provides documentation of the extent of competency through early identification of deficiencies and help in overcoming them.

5. Places the responsibility for improvement with the employee thereby raising level of concern and desire to improve.

The Redfern Model, like any approach, is open to questions related to its underlying assumptions and application of technical concepts. We have forsaken the latter to concentrate on the propositions which anchor the system. We raise three questions; the first is related to what Redfern has labelled "a partnership in evaluation," the second deals with the type of written objectives used, and the third addresses the use of student achievement scores as an evaluation criterion.

Question One: Is performance evaluation a partnership?

This is a most important question, for mutuality is the cornerstone upon which the model is built. Redfern (1980) maintains that evaluation and supervision are compatible, that a performance objective approach removes the hesitancy that supervisors generally exhibit toward getting actively involved in a rating type of program, and that it encourages a partnership between the evaluator and evaluatee. These assumptions are music to the optimists' ears but discordant to the pragmatist. Given the preciousness of gainful employment in today's labor market, the frailties of human nature, and the tenuous relationship between principals and teachers, one must wonder if mutual goal-setting is possible. The difficulties inherent in the self-formulation of realistic job targets has been well documented in times of less stress and strain (NSPRA, 1974). One wonders about the equality of the partnership between evaluator and evaluatee. In his latest book, Evaluating Teachers and Administrators; A Performance Objective Approval, (1980) the author speaks of "partners in the duet" in describing responsibilities in carrying out the actual plan. It almost sounds like a discussion of clinical supervision. Performance evaluation and clinical supervision should not be confused. Evaluation is designed to assess competence, despite whatever other lofty goals if may aspire to. Clinical supervision is concerned with professional growth not inspection.

There may be no inherent danger in packaging a model as a partnership: indeed, it is less threatening to participants. Those who must implement it however, will be unprepared and rudely awakened when problems surface.

Question Two: Can expected outcomes be effectively translated into behavioral terms?

This question has been raised before. Cardillechio (NASSP Bulletin, 1974) explored the issue nearly a decade ago. Written performance objectives are a vital component of the model. Redfern says that there is a place for general goals, but they are difficult to measure and don't lend themselves to documentation. He lauds the elegance of precise behavioral objectives. Few would disagree that behavioral objectives simplify outcomes in such a way that learning is weakened, but by Redfern's own admission they are of little use in formulating targets in the affective domain. While Redfern advises prospective users to ponder the merits and demerits of both general goals and behavioral objectives, he provides little guidance as to what they should base their decisions on. In his words, "time and circumstances will determine the best approach". What circumstances should be considered and how time is related to formulating what type of objectives are important but unanswered questions.

Question Three: Is it valid or even fair to assess performance outcomes in terms of student gains?

Few researchers would give this question an unqualified yes. It appears to be overly simplistic and potentially dangerous to mix individual employee evaluation with student outcomes, without rigorous control for extraneous factors. There is no evidence that cultural and school evironments, budget contraints and administrative leadership have been considered in formulating objectives and one doubts if it is within the realm of possibility to do so. Deletion of this aspect of the model seems advisable.

The TPE Model

TPE is a product of many years of research, investigation, and practical experience in assessing teacher competency.[2] It flows from two basic assumptions: one, participatory planning offers maximum opportunity to incorporate organizational members' preferences and needs into the system, and two, formative and summative evaluations are both necessary components of a successful evaluation system.

Planning

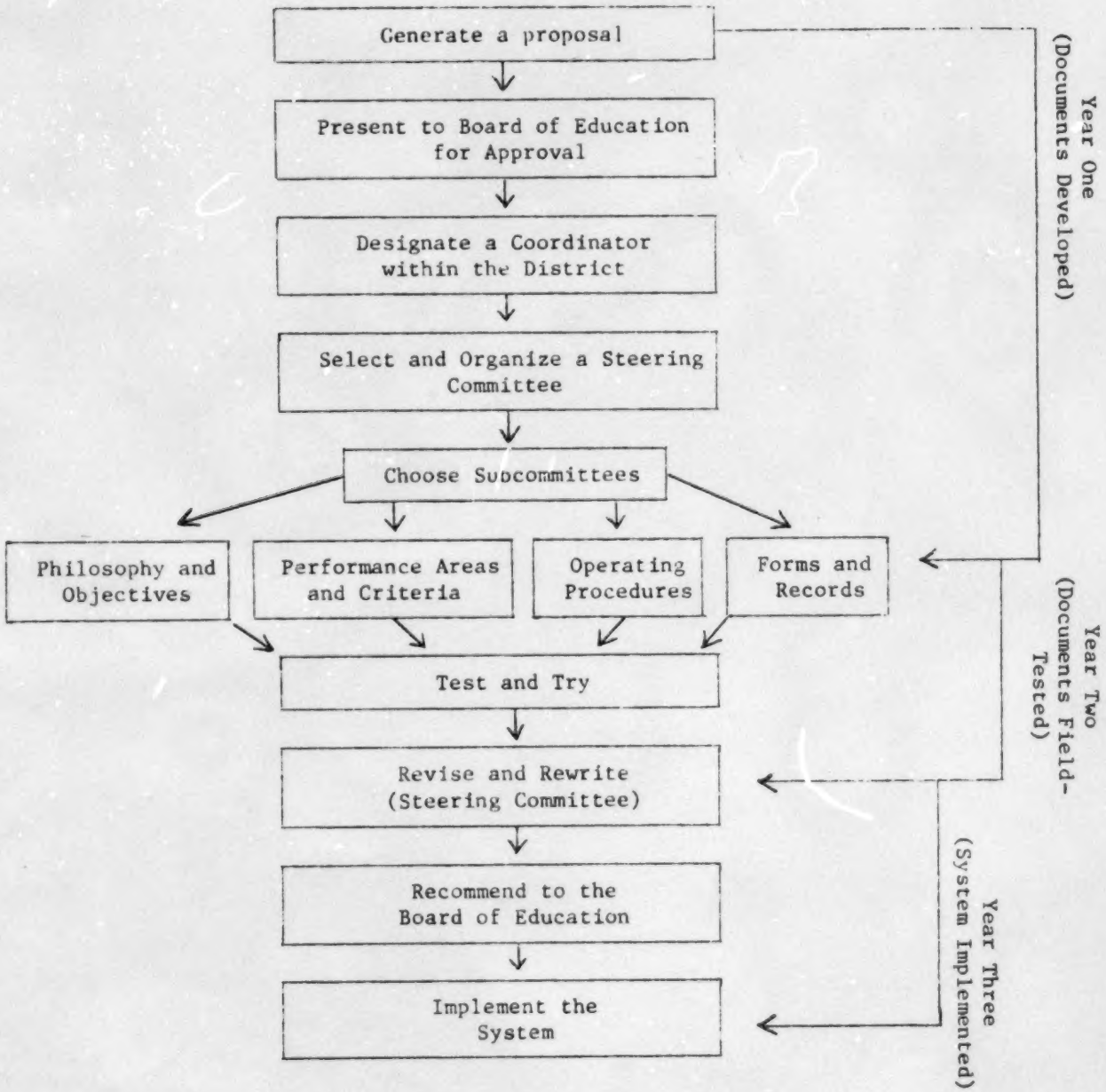
Planning is a prerequisite for most organizational endeavors. TPE enables the school organization to involve the school community in planning a system which is in concert with the school's goals and philosophy, while providing an opportunity to test and try the system. The planning process includes development of three major components: (1) evaluation instruments for teachers, (2) evaluator training, and (3) ongoing staff development to improve instructional leadership. Figure 2 shows the three-year time frame, as well as the system components.

The steering committee serves as the fulcrum of the process. Ten to twenty people are chosen according to the contribution each can make to the total effort. Typically, the committee includes teachers from curriculum areas and grade levels, administrators, parents, board of education members, and secondary school students. The committee's major tasks are to: (1) create and manage the timeline, (2) serve as a link to the board and superintendent, (3) determine the use of consultants, (4) guide system development, and (5) inform and consult with the staff.

---

[2] TPE was developed by Dr. Richard Manatt, Professor of Education, E005 Quadrangle, Iowa State University.

Figure 2. Time Frame for Developing a Performance Evaluation System


Generate a proposal
Present to Board of Education for Approval
Designate a Coordinator within the District
Select and Organize a Steering Committee
Choose Subcommittees
Philosophy and Objectives
Performance Areas and Criteria
Operating Procedures
Forms and Records
Test and Try
Revise and Rewrite (Steering Committee)
Recommend to the Board of Education
Implement the System
Year One (Documents Developed)
Year Two (Documents Field-Tested)
Year Three (System Implemented)

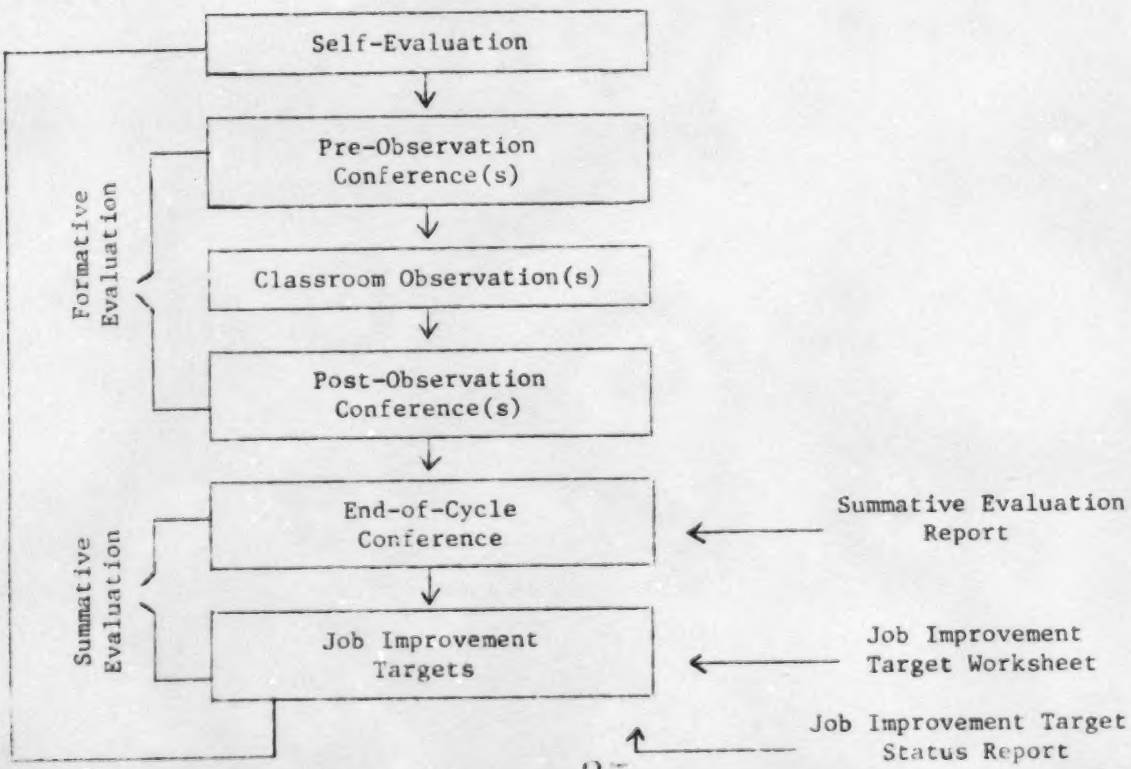
Usually, five subcommittees are designated each with a specific charge. As they meet and confer, they generate ideas that are discussed with the total

group and used to develop prototypic instruments and procedures. The Philosophy and Objectives Subcommittee has an important role; they must clearly define the purpose of the evaluation system and communicate this across the school organization. Productive teaching behavior is defined by the Performance Areas and Criteria Subcommittee while the Operational Procedures Subcommittee determines procedures and establishes the TPE cycle. The Form and Research Subcommittee determines the most efficient way to record data and developes the appropriate documents.

The TPE Cycle

The TPE cycle is an integral part of the model. Figure 3 shows the flow of activities which comprise a cycle which provides both formative and sum-

Figure 3. Flow Chart of a TPE Cycle



mative evaluation. The pre-observation conference begins the formative process. The teacher and evaluator meet briefly and discuss the lesson to be observed. The evaluator then observes the lesson using whatever data gathering techniques are appropriate, considering the objectives of the lesson as well as other factors. Since the data gathered in the lesson is for formative puposes, it is usually recorded in narrative form although it may be reflected in frequencies of teacher or learner behaviors. These data are used in the post-observation conference as the basis for repeated discussion of the teacher's performance in the lesson. This process may be repeated several times, depending on contractual agreements and need.

The end-of-cycle conference is a summative assessment of the teacher's performance. The assessment is recorded on the summative evaluation instrument which was developed by the Forms and Records Subcommittee. During the end-of-cycle conference, job improvement targets are discussed and finalized in writing soon after. The Summative Evaluation Report and job improvement targets are used to initiate reentry into the cycle the following year.

The Test-and-Try Subcommittee, in addition to planning the orientation and determining the extent of training needed by the evaluators, designs the field test that precedes system implementation. While TPE is a stand-alone system, many school organizations tie Administrator Performance Evaluation (APE) to it, implementing TPE in the first year and APE in the second.

Bouquets and Brickbats

There are two sources of evidence for the analysis of this model. The first is application of the criteria, the second, a data base generated by its track record in a number of school systems. TPE has the following advantages:

1. It provides an opportunity for the participants to tailor the system to meet their needs.
2. It provides a timeline and methods for field testing procedures and prototypic instruments as well as for refining both.
3. The steering committee and subcommittees provide optimum opportunity for teacher and administrator interaction, involvement and consultation.
4. The TPE cycle provides for both formative and summative evaluation. It guarantees teachers the opportunity to receive assistance throughout the year while guaranteeing the board of education and community a measure of competence that is valid, reliable, and legally discriminating.
5. Job improvement targets define the professional development goals of teachers, the timeline for reaching more goals, and the role of the supervisor in assisting teachers to achieve their goals.
6. It provides for professional growth activities for evaluatees and evaluators.

In summary, TPE contains the essentials of an effective system: a logical and workable sequence, procedures, and instrumentation that fit the organization, a mechanism to enhance consultation and information giving, opportunity for continuous performace improvement, and components which are conceptually sound and congruent with the organization, philosophy and objectives.

TPE is not a panacea, nor has it achieved perfection. While one of its obvious strengths is participatory planning, this also is open to question. Incorporating a clinical supervision cycle into a process which culminates in summative evaluation demands discussion. Finally, there is the matter of job improvement targets. Three questions address these issues.

Question One: Should the planning process be revised?

The TPE planning process explicitly provides for representative participation in decisions regarding system development and implementation. Yet, because of its extensive usage, TPE has been refined to a point where many of its instruments and processes are highly valid and reliable. While there are

still a number of important decisions for the steering committee and subcommittees to make, the leadership is often more intent on gaining acceptance of the system. For example, the Philosophy and Objectives Subcommittee, has the very real task of developing and articulating the purpose of the system but some other subcommittees act more as checkers, making small revisions in existing documents or adjusting time periods. There may be a danger in this. As teachers' organizations become more concerned and involved in the evaluation process, perhaps it would be wise to clarify the extent of the discretion which subcommittees have in decisionmaking.

Question Two: Should job improvement targets be included in the formative cycle?

In a number of school organizations the busy principal observes the teacher once and holds a post-observation conference which is also the end-of-cycle conference. This short-circuits the cycle, and is of little assistance in helping the teacher improve. It seems likely that if it is obligatory for the supervisor to conduct a classroom observation early in the year, then assist the teacher in developing job improvement targets, and later observe the teacher and conduct a post-conference, it would enhance growth and make the end-of-cycle conference and end-of-year job improvement target more salient. While it is presently recommended that supervisors hold two or more conferences with teachers, many principals, claiming a lack of time, revert to the minimum contractual stipluation. Building it into the system might help to resolve that problem.

Question Three: Can TPE improve on the use of job improvement targets?

Despite the fact that principals and supervisors receive vigorous train-

ing in writing job improvement targets, analysis reveals that they tend to be written in unimportant areas and do not meet criteria for effectiveness. They are often vague and fail to specify what behaviors the teacher must exhibit to meet the target. A bank of important job improvement targets, written in a manner so as to make them an effective vehicle for improvement, might alleviate this weakness. Obviously, the principals must add to them specifying target data and the assistance to be provided.

Futures

Our world is a new world. The very difficulties which it presents derive from growth in understanding, in skill, and in our ability to use the resources which reside within us, both in an ideographic and normative sense. The future of teacher evaluation rests on our ability to transcend our myopia. Teacher assessment is here and it's here to stay. One is reminded of the old parable involving the grasshopper and the owl. It seemed that the grasshopper suffered from severe pains in the winter due to the savage temperature. In distress, he sought the advice of the wise old owl who gave him this counsel, "Simply turn yourself into a cricket and hibernate during the winter." Having received this wonderful advice the grasshopper was overjoyed. Joy turned to distress, however, when he discovered this knowledge could not be transformed into action. Upon confronting the owl he was admonished, "Look, said the owl, I gave you the principle. It's up to you to work out the details." The moral has value to those who assess teacher competence. We must desist from tilting at windmills, build our knowledge base, and work out the details.

There is a desperate need to develop teacher evaluation centers at se-

lected sites across the nation. While institutions of higher education should take the lead, the excellent work done by other agencies suggests they can aid greatly in the effort. These centers should take dead aim at validating and refining instruments and procedures. This should be a cooperative venture between school organizations and centers, and between centers. In addition to conducting research and developing the system, centers should provide assistance to school organizations and disseminate important findings and advancements to them.

The next question to be addressed seems obvious: In what specific areas should we concentrate the research and development effort? Our extensive literature review and experience with TPE suggest that we focus on the following:

1. Continue and broaden the effort to identify effective teaching criteria. Despite the tremendous value of recent research findings we must go further, by investigating the relationship between situational elements and effective teaching. We also need to identify criteria for effectiveness in special areas (e.g., music, practical arts, physical education, etc.).
2. Perform task analysis on each of the evaluation components. Classroom observation and conducting supervisory conferences should be high on that list for data gathering and feedback are the heart of evaluation.
3. Develop training packages for teaching evaluators necessary skills. No model is efficacious unless it provides for effective training to evaluators. As classroom observation and supervisory conferences are the heart of evaluation, evaluator training is its artery.
4. Develop a subsystem for assisting the marginal or unsatisfactory teacher. There is little substantive information on how to help these individuals. It is a subject of great concern to evaluators.
5. Develop efficient ways to teach teacher evaluators the elements of effective teaching. Approaches which consume in excess of 500 clock hours are admirable, but consume time and money--two resources in short supply. It seems likely that concepts and methods could be condensed and implemented more efficiently.

We have omitted a dimension which has tremendous potential -- the use of the computer in evaluation research and development. The possibilities for

creating data banks and using them in many ways are unlimited. One would hope that the power of the computer in research and development will not be overlooked.

Final Thoughts

The most pervasive aspect of teacher evaluation is its inevitability. Describing the state of the art is bittersweet; the lack of uniformity is disturbing but progress made in the last half-decade is encouraging. Earlier in the paper it was pointed out that teacher assessment does not occur in a vacuum, it has a context, tethered to social action and thought. It remains an open question whether personnel evaluation, which is such a highly individual business, can flower in today's social milieu. It is a practical imperative that we give to this effort every possible resource that we can muster. If we are going to make a difference in the lives of our young people it is virtually guaranteed that their classrooms and teachers will play a significant role. The challenge to educators is to wed theory and practice and enrich both to the extent that they improve and guarantee teacher competence. It seems worth the effort.

References

Acheson, K., and Gall, M. Techniques in the Clinical Supervision of Teachers. New York: Longman, 1980.

Aleamani, L. M. Student ratings of instruction. In J. Millman (ed.), Handbook of Teacher Evaluation. Beverly Hills, CA: Sage Publication, 1981, 110-145.

Blumberg, A. Supervision and Teachers: A Private Cold War. Berkeley, CA: McCutchan, 1980.

Bobbitt, J. F. The elimination of waste in education. The Elementary School Teacher, 12, 260.

Bolton, D. Selection and Evaluation of Teachers. Berkeley, CA: McCutchan, 1973.

Borich, G. D. The Appraisal of Teaching: Concepts and Process. Reading, MA: Addison-Wesley, 1977.

Boyce, A. C. Methods of measuring teachers' efficiency. 14th Yearbook, National Society for the Study of Education, Part 2. Chicago: University of Chicago Press, 1915.

Brandt, R. M. Toward a taxonomy of observational information. In Observational Methods in the Classroom, pp. 23-34. Edited by Charles W. Beegle and Richard M. Brandt. Washington, D.C.: Association for Supervision and Curriculum Development, 1973.

Bridges, E. M. Faculty evaluation--A critique and a proposal. Administrator's Notebook, 1974, 22, 1-4.

Cardellichio, T. L. Evaluating teachers' methods. NASSP Bulletin, December 1974, 58, 8-12.

Cubberly, E. Public School Administration. Boston, MA: Houghton Mifflin, 1916.

Davis, H. Evolution of current practices in evaluating teacher competence. In B. J. Biddle & W. J. Ellena (Eds.), Contemporary Research on Teacher Affectiveness. New York: Holt, Rinehart and Winston, 41-67.

Evaluating Teacher Performance (ERS Report), Arlington, VA: 1978.

Flanders, N. A. "Interaction Analysis in the Classroom: A Manual for Observors." Unpublished manuscript, University of Michigan, 1960.

Galloway, C. M. The nonverbal realities of classroom life. In C. Beegle & R. Brandt (Eds.), Observational Methods in the Classroom. Washington, D.C.: Association for Supervision and Curriculum Development, 1973, 44-45.

Glass, G. V. A review of three methods of determining teacher effectiveness. In H. J. Walberg (Ed.), Evaluating Educational Performance. Berkeley, CA: McCutchan, 1974, 11-32.

Griffith, F. A Handbook for the Observation of Teaching and Learning. Midland, MI: Pendell Publishing Company, 1973.

Hoole, C. Scholastic discipline. American Journal of Education, (Originally published in 1659), 17, 293-324.

King, L. A. The present status of teacher rating. American School Board Journal, 1925, 70, 44-46.

Lewis, A. C. Evaluating Educational Personnel. Arlington: American Association of School Administrators, 1982.

Mayo, E. The Social Problems of an Industrial Civilization. Cambridge, CA: Harvard University Press, 1933.

Medley, D. M. Measuring the complex classroom of today. In Observational Methods in the Classroom, edited by Charles W. Beegle and Richard M. Brandt. Washington, D.C.: Association for Supervision and Curriculum Development, 1973, 35-44.

Millman, J. Handbook of Teacher Evaluation. Beverly Hills, CA: Sage Publications, 1981.

National School Public Relations Association. Evaluating Teachers for Professional Growth: Current Trends in School Policies and Programs. Arlington, VA, 1974, 64.

Pine, G. J., & Boy, A. V. Necessary conditions for evaluating teachers. NASSP Bulletin, 1975, 59, 18-23.

Popham, W. J. & McNeil, J. D. The assessment of teacher competence. In R. M. Travers (Ed.), Second Handbook of Research on Teaching. Chicago: Rand McNally, 1973, 230-236.

Reavis, W. C., & Cooper, D. H. Evaluation of Teacher Merit in City School Systems. Supplementary Educational Monographs, Number 59. Chicago, IL: University of Chicago, 1955.

Redfern, G. B. Evaluating Teachers and Administrators: A Performance Objectives Approach. Boulder, CO: Westview Press, Inc., 1980.

Roethlisberger, F. J., & Dixon, W. J. Management and the Worker. Cambridge, MA: Harvard University Press, 1939.

Stow, S., & Sweeney, J. Developing a teacher performance evaluation system. Educational Leadership, April 1981, 38:7, 538-542.

Strike, K., & Bull, B. Fairness and the legal context of teacher evaluation. In J. Millman (Ed.), Handbook of Teacher Evaluation. Beverly Hills, CA: Sage Publication, 1981, 303-341.

Sweeney, J., & Stow, S. Performance improvement--A people program. Education, Spring 1981, 101:3, 267-269.

Taylor, F. W. The Principles of Scientific Management. New York: Harper and Row, 1911.

Willes, W., & Lovell, J. T. Supervision for Better Schools. Englewood Cliffs, NJ: Prentice-Hall, Inc., 1975.

Wuhs, S. State Mandates For Evaluation. Unpublished Manuscript, 1982.

APPENDIX A

States Requiring Teacher Evaluation

States	Statute/ Regulation	Year(s) Established	Purpose		
			Dismissal	Improve-ment	Account-ability
Alabama					
Alaska					
Arizona	Statute	1981		x	
Arkansas	Statute	1979	x		
California	Statute	1972/1976		x	
Colorado					
Connecticut	Statute	1974		x	
Delaware					
Florida	Statute	1961		x	
Georgia*					
Hawaii	Statute	1965	x	x	
Idaho					
Illinois					
Indiana					
Iowa	Statute	1976	x		
Kansas	Statute	1973		x	
Kentucky	Regulation	1981		x	
Louisiana	Statute	1977/1980		x	
Maine					
Maryland					
Massachusetts					
Michigan					
Minnesota					
Mississippi					
Missouri					
Montana					
Nebraska					
Nevada	Statute	1973/1979	x	x	
New Hampshire					
New Jersey	Both	1973/1977/ 1979		x	
New Mexico	Regulation	1974/1981		x	
New York					
North Carolina	Statute	1982		x	x
North Dakota					
Ohio	Both	1968/1980		x	
Oklahoma	Statute	1977	x	x	
Oregon	Statute	1971/1979	x	x	
Pennsylvania	Both	1949 (as amended)	x	x	
Rhode Island					
South Carolina	Statute	1979	x	x	
South Dakota	Statute	1969/1975	x	x	
Tennessee	Statute	1979		x	
Texas		1975	x	x	
Utah					
Vermont	Regulation	1981			x
Virginia	Statute	1972		x	
Washington	Statute	1975	x	x	
West Virginia	Statute	Unknown	x	x	
Wisconsin					
Wyoming					

*Georgia requires teacher performance appraisal for probationary teachers only (3 years).

Appendix B

Evaluation System Criteria

1. Response mode of instrument.

2. Is there a lesson assessment required?

3. Is a preconference provided?

4. Are classroom observations called for?

5. Is a postconference provided?

6. Are formative reports written each time an observation cycle is conducted?

7. Is an annual summative report provided?

8. Does clinical supervision play a part of the cycle?

9. Are there comparisons made with standards of teacher performance?

10. Is intensive assistance provided?

11. Is there provision for recognition of outstanding teacher performance?

12. Are job improvement targets used?

13. Does the cycle provide procedural and substantive due process?

14. Can the performance appraisal standards be communicated clearly as rules of the game?

15. Is it useful for evaluating nontenured teachers, tenured teachers, and for use for transfer, promotion, and job assignment decisions?

16. Does the instrument have provision for self-evaluation? Peer evaluation? Administrator evaluation? Student evaluation? Achievement testing?

17. Do the operational procedures provide for "stretching and stroking" the super teacher as well as helping the poor one?

18. Are links made to sequence and scope of the curriculum?

19. Does the system provide help beyond the principal?

20. Does the procedure hook to staff development activities?

21. Is the evaluation system too costly?

22. Does the evaluation system avoid being "too labor intensive" for the principal?

23. Are work samples taken?

24. Is the cycle clear?

25. Is the initial assessment based on observation, not a checklist?

26. Are the assessors/appraisers rigorously trained?

27. Is interrater reliability checked periodically?

28. Is the cycle short enough to provide data for management and Board decisions annually?

29. Are there only 15-20 criteria?

30. Are the response modes descriptive (e.g., semantic differential or graphic response mode?)?

31. Are there multiple responses (not just satisfactory/unsatisfactory, five choices or ten choices are better)?

32. Are the criteria weighted for importance?